Learning effective color features for content based image retrieval in dermatology

Bunte, Kerstin; Biehl, Michael; Jonkman, Marcel F.; Petkov, Nicolai

Published in:
Pattern recognition

DOI:
10.1016/j.patcog.2010.10.024

Publication date:
2011

Pattern Recognition 44 (2011) 1892–1902

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Learning effective color features for content based image retrieval in dermatology

Kerstin Bunte [a,*], Michael Biehl [a], Marcel F. Jonkman [b], Nicolai Petkov [a]

[a] Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, The Netherlands
[b] Department of Dermatology, University Medical Center Groningen, University of Groningen, The Netherlands

ARTICLE INFO

Available online 7 November 2010

Keywords:
Machine learning
Learning vector quantization
Adaptive distance measures
Content based image retrieval

ABSTRACT

We investigate the extraction of effective color features for a content-based image retrieval (CBIR) application in dermatology. Effectiveness is measured by the rate of correct retrieval of images from four color classes of skin lesions. We employ and compare two different methods to learn favorable feature representations for this special application: limited rank matrix learning vector quantization (LiRaM LVQ) and a Large Margin Nearest Neighbor (LMNN) approach. Both methods use labeled training data and provide a discriminant linear transformation of the original features, potentially to a lower dimensional space. The extracted color features are used to retrieve images from a database by a k-nearest neighbor search. We perform a comparison of retrieval rates achieved with extracted and original features for eight different standard color spaces. We achieved significant improvements in every examined color space. The increase of the mean correct retrieval rate lies between 10% and 27% in the range of $k=1$–25 retrieved images, and the correct retrieval rate lies between 84% and 64%. We present explicit combinations of RGB and CIE-Lab color features corresponding to healthy and lesion skin. LiRaM LVQ and the computationally more expensive LMNN give comparable results for large values of the method parameter κ of LMNN ($\kappa \geq 25$) while LiRaM LVQ outperforms LMNN for smaller values of κ. We conclude that feature extraction by LiRaM LVQ leads to considerable improvement in color-based retrieval of dermatologic images.

1. Introduction

In the last decades the availability of digital images produced by scientific, educational, medical, industrial and other applications has increased dramatically. Thus, the management of the expanding visual information has become a challenging task. Since the 1990s Content Based Image Retrieval (CBIR) is a rapidly advancing research area, which uses visual content to search images from large databases according to the user's interest [36,23,21,12,22,26,41,17]. A typical CBIR system extracts visual information from an image and converts it internally to a multidimensional feature vector representation. For retrieval, the dissimilarities (distances) between the feature vector of a query image and the feature vectors of the images in the database are computed. Then, the database images most similar to the query are presented to the user. CBIR may especially be interesting in the field of computer aided diagnostics when it is partly based on images. An intelligent pre-selection of images with a trained system might help a medical doctor to efficiently search for patients, who had problems similar to the actual case.

The visual content of an image can be described by color, texture, shape or spatial relationship. A good visual content descriptor should be insensitive to the specific imaging process, e.g. invariant under changes of illumination. The prevalent visual content for image retrieval is color. Frequently used color descriptors are color moments, histograms, coherence vectors and correlograms [33,24]. Before a color descriptor can be selected, the underlying color space has to be specified.

There are many different color spaces available, which may be beneficial in different application domains. The color representations most commonly used in electronic systems are RGB and CIE-XYZ. CIE-XYZ and the related CIE-Lab and CIE-Luv are designed to match human perception. In [40] the authors argue, that normalized TSL (Tint, Saturation, Lightness) is superior to other color spaces for skin modeling with a unimodal Gaussian joint probability density function. The color space YCrCb is adjusted for efficient image compression, but the transformation simplicity and explicit separation of luminance and chrominance components appear attractive for skin color modeling [25,46,9]. Surveys on color spaces and their use can be found in [40,43]. We are not aware of a general rule for the choice of the color space and the representation might follow the users preference. So we decided to investigate eight different color spaces, which are commonly used and may be useful for the task at hand.

* Corresponding author. Tel.: +31 50 3637049; fax: +31 50 3633800.
E-mail addresses: k.bunte@rug.nl (K. Bunte), m.biehl@rug.nl (M. Biehl), m.f.jonkman@derm.azg.nl (M.F. Jonkman), n.petkov@rug.nl (N. Petkov).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.10.024

Color is an important attribute for primary skin efflorescences [3]. Color features have proven beneficial in many applications and medical sciences, especially for the recognition of skin lesions [14,40,43,38,34,19,37,25,46,18] or the classification of skin cancer [28,44,1,16,10,42]. A dermatologist might be interested in pictures of similar skin lesions in comparison to an actual case to verify the diagnosis or confer with similar symptoms. This can be interpreted as a problem of CBIR. The authors of [4] study the use of color features and the effectiveness of different color spaces in this context. They conclude that the representation of an image by the difference in the average color of healthy and lesion skin gives better results than the explicit use of the pair of colors. Fig. 1 shows two example retrievals for a CBIR system in the field of skin lesion comparison in dermatology. In [4], the best results were achieved with the CIE-Lab color representation.

Since the difference of two color values is a special case of a linear transformation, the question arises whether better results can be achieved by more general linear transformations. Of course, it is possible that the use of a combination of a cyclic distance measure in the case of color spaces containing a "hue"-descriptor might lead to superior results. We will address this interesting questions in further studies. One well known technique to achieve a linear projection of feature vectors to a subspace which minimizes the overlap between different classes is Linear Discriminant Analysis (LDA) [13]. In this paper we employed and compared two different recent techniques, which are able to find discriminant feature transformations based on a supervised training procedure. The Large Margin Nearest Neighbor (LMNN) [45] approach has the advantage that it is based on a convex cost function, so it returns the global optimum for the current configuration of training data and parameters, based on the kNN approach. The Limited Rank Matrix Learning Vector Quantization (LiRaM LVQ) [29,30,8,32] on the other hand follows a stochastic gradient descent procedure and may get stuck in local minima. On the other hand, it has the advantage of low computational costs. It is a prototype-based method, in which the decision boundary is defined by the Voronoi cells of prototypes following the large margin principle [11]. Both algorithms are available in general form and turned out to be effective classifiers in many applications. In our real world example application of CBIR in dermatology, the LiRaM LVQ approach turned out to be quite robust concerning the initialization and parameter setting. With comparably low computational costs it leads to similar or better results than the LMNN approach with optimal parameter setting on most color spaces discovered. We improve the correct retrieval rate in CBIR of dermatological images significantly by applying adaptive linear transformations.

The main aim of this work is to demonstrate in terms of a real world example, that an adaptive, i.e. data driven transformation of original color features can improve the retrieval performance of a CBIR system significantly. We concentrate on the performance enhancement achieved by using the most basic, easy and fast acquirable set of important features for the problem at hand, i.e. color information only.

In Section 2 we explain the real world data set, the feature extraction process, we present and discuss the methods we use to determine optimal transformations of color features and their use in the CBIR system. In Section 3 we present results and conclude in Section 4.

2. Methodology

An illustration of the Methodology is shown in Fig. 2.

2.1. Data set and feature extraction

We analyze images from a database maintained at the Department of Dermatology of the University of Groningen. At the time of this study it consisted of 47,621 images from 11,361 patient sessions, the number of images grows by about 5000 per year. Clinical images are obtained under standard light conditions and do not require further calibration. A subset of 211 images was provided and manually labeled by a dermatologist, who assigned each image to one of four classes of lesions. For better readability we refer to these classes as "red", "white", "blue" and "brown" (Fig. 3). These terms correspond to the relative tint of lesions which appear reddish, blue, brownish or hypopigmented on the background of the surrounding healthy skin. We consider a data set with 82, 46, 29 and 54 samples, respectively, which amounts to a total of 211 images.

Of course there are more characteristics then just color which identify the kind of skin lesion, e.g. the shape. The consideration of other types of features will be addressed in future work, here we concentrate on the quality the most basic set of features is able to achieve. In this particular problem color seems to be a suitable indicator for the skin lesion classes. The complete data set also contains other skin lesions, but in this study we restrict ourselves to the consideration of the above mentioned classes. Here, emphasis is not on the classification performance itself. It serves as a basis for improving the retrieval system and the supervised training yields a suitable distance measure. Further studies should address additional features, more general skin lesion classes and the handling of unknown classes.

The original images were not pre-processed. For each image a region of lesion and a region of healthy skin are manually selected and for each of them the average color values are computed (see Fig. 4). Hence, the extracted data contains three color components for each of the two regions, resulting in a six-dimensional (6D) feature vector. As a normalization step we perform a z-score-transformation resulting in zero mean and unit variance features. This normalization is reasonable in the RGB color space and linear domains. In case of cyclic descriptors, like the "hue", this might not be appropriate. The combination of cyclic distances and linear dissimilarities and their normalization concerning this specific task will be addressed in future studies. Nevertheless, for the sake of comparison and completeness we show the results on different color spaces under the same conditions.

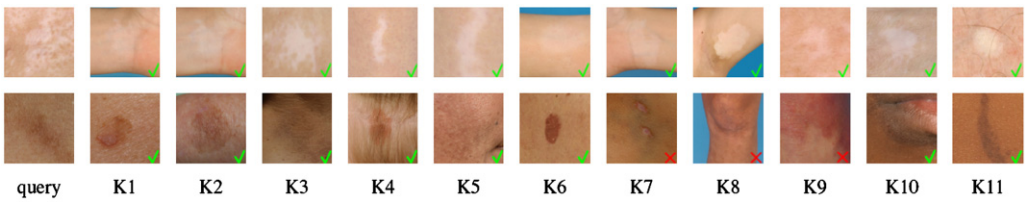

Fig. 1. Two example retrievals of the 11 most similar images for a given query image. The first image in a row is the query image, followed by the images returned from the retrieval system [4]. The green tick marks images with the same class label like the query.

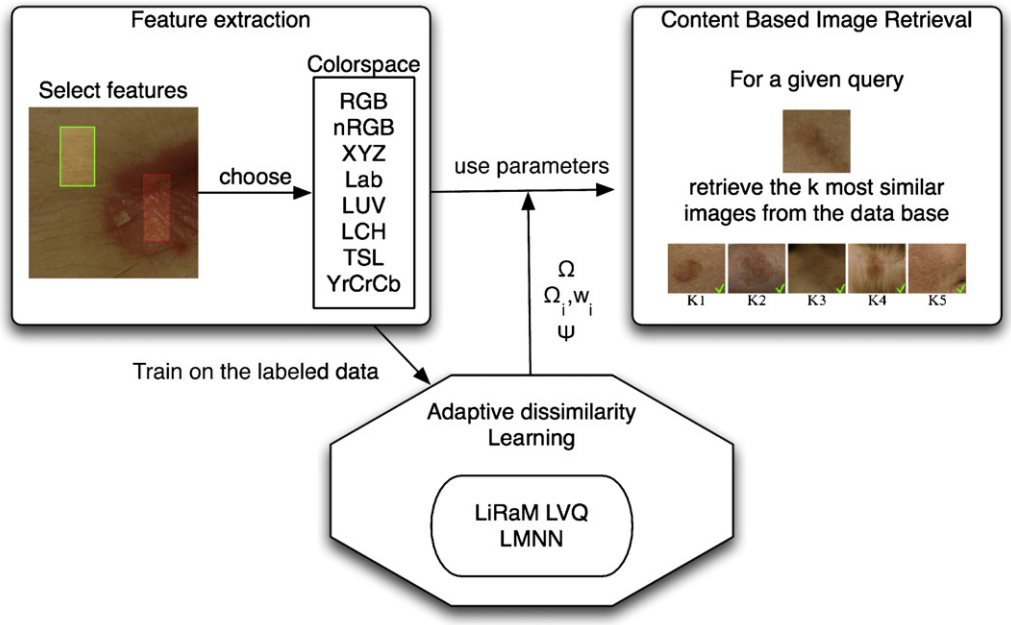

Fig. 2. Methodology overview for the proposed CBIR system.

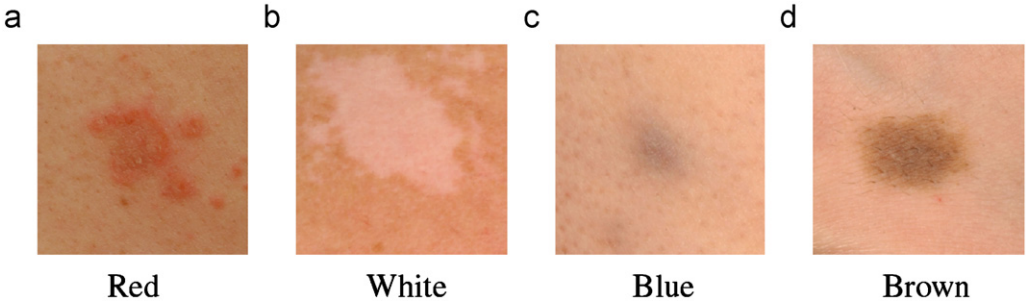

Fig. 3. Example images of the four skin lesion classes taken from [4].

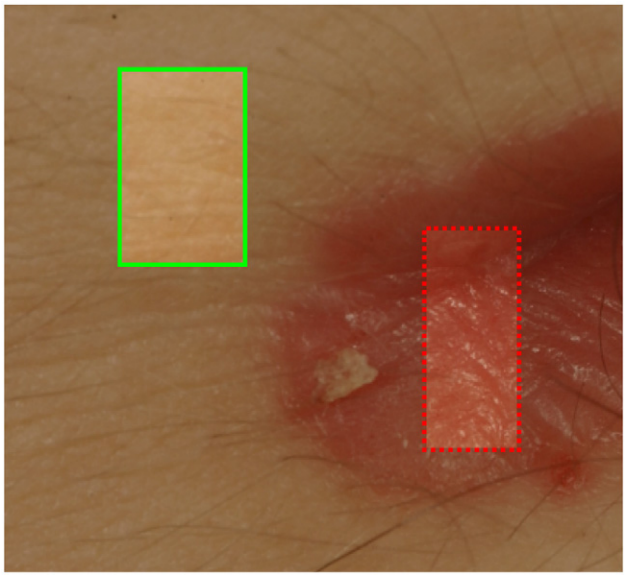

Fig. 4. Feature extraction (taken from [4]): a representative region of healthy skin (green framed) and lesion skin (red framed) were manually selected. The average colors of these two regions are combined in a six-dimensional feature vector.

2.2. Feature transformation obtained by LiRaM LVQ

In order to obtain discriminative representations of the data we use supervised machine learning. Specifically, we employ limited rank matrix learning vector quantization (LiRaM LVQ), a recently introduced method which adapts a similarity or distance measure in the course of learning [6,8,32]. It is an extension of Generalized LVQ (GLVQ), which is a prototype based classification algorithm and a modification of Kohonen's heuristic LVQ [20]. These methods aim at the quantization of the data space in form of a finite number of prototypes defined in the same feature space as the data. At the end of the learning process the prototypes may be interpreted as typical representatives of the given classes. These methods exhibit, among other things, the advantage of easy interpretation, implementation and model flexibility. Recently, a number of extensions has been introduced. GLVQ updates prototypes by means of gradient descent with respect to a heuristically motivated cost function suggested by Sato and Yamada [27]:

$$f_C = \sum_i \Phi(\mu) = \sum_i \Phi\left(\frac{d_J^A - d_K^A}{d_J^A + d_K^A}\right). \quad (1)$$

Here, Φ is a monotonic function, e.g. the logistic function or the identity $\Phi(x) = x$ which we will consider throughout the following.

Generalized Matrix LVQ (GMLVQ) takes into account the importance of single features as well as pairwise correlations between them by means of a full matrix Λ of relevances [29,30]. In addition, LiRaM LVQ limits the rank of the relevance matrix to obtain transformations into a low-dimensional space [6,8]. Training is based on examples of the form $(\boldsymbol{\xi}_i, y_i) \in \mathbb{R}^N \times \{1,\ldots,C\}$, where N is the dimension of feature vectors and C is the number of classes (in our case, $N=6$ and $C=4$). At least C prototypes, which are chosen as typical representatives of the respective classes, are characterized by their location in feature space $\boldsymbol{w}_i \in \mathbb{R}^N$ and the respective class label $c(\boldsymbol{w}_i) \in \{1,\ldots,C\}$. Given a parameterized distance measure $d^{\Lambda}(\boldsymbol{w},\boldsymbol{\xi}) \in \mathbb{R}$, the classification is performed according to a "winner takes all" or "nearest prototype" scheme: A data point $\boldsymbol{\xi} \in \mathbb{R}^N$ is assigned to the class label $c(\boldsymbol{w}_i)$ of the closest prototype i with $d^{\Lambda}(\boldsymbol{w}_i,\boldsymbol{\xi}) \leq d^{\Lambda}(\boldsymbol{w}_j,\boldsymbol{\xi}) \forall j \neq i$.

Learning is an iterative procedure which presents a single example at a time (step) and moves prototypes closer to (away from) data points representing the same (a different) class. In every step also the distance measure is modified, usually with a different (smaller) learning rate [30]. It is parameterized by an adaptive matrix $\Lambda \in \mathbb{R}^{N\times N}$, which can account for correlations between different features:

$$d^{\Lambda}(\boldsymbol{w},\boldsymbol{\xi}) = (\boldsymbol{\xi}-\boldsymbol{w})^{\top}\Lambda(\boldsymbol{\xi}-\boldsymbol{w}). \tag{2}$$

This can be seen as a generalization of the well known Euclidean distance including mixing of attributes. Since the matrix Λ is assumed to be positive (semi-) definite, the measure corresponds to the (squared) Euclidean distance in an appropriately transformed space: $\Lambda = \Omega^{\top}\Omega$ and, hence, $d^{\Lambda}(\boldsymbol{w},\boldsymbol{\xi}) = [\Omega(\boldsymbol{\xi}-\boldsymbol{w})]^2$. In [8,32] the formalism has been extended to the use of rectangular matrices Ω, which define transformations from the original N-dimensional feature space to $\mathbb{R}^M$ with $M \leq N$. The corresponding algorithm is referred to as LiRaM LVQ.

In order to formulate stochastic gradient descent with respect to the objective function (1) we compute the derivatives

$$\frac{\partial d_L^{\Lambda}}{\partial \boldsymbol{w}_L} = -2\Omega^{\top}\Omega(\boldsymbol{\xi}-\boldsymbol{w}_L) = -2\Lambda(\boldsymbol{\xi}-\boldsymbol{w}_L), \tag{3}$$

$$\gamma^{+} = \frac{\partial \mu}{\partial d_J^{\Lambda}} = \frac{2d_K^{\Lambda}}{(d_J^{\Lambda}+d_K^{\Lambda})^2}, \tag{4}$$

and

$$\gamma^{-} = \frac{\partial \mu}{\partial d_K^{\Lambda}} = \frac{-2d_J^{\Lambda}}{(d_J^{\Lambda}+d_K^{\Lambda})^2}. \tag{5}$$

Here, $L \in \{J,K\}$ and the index J (K) refers to the closest correct (wrong) prototype $\boldsymbol{w}_J$ ($\boldsymbol{w}_K$).

For the closest correct prototype $\boldsymbol{w}_J$ and closest wrong prototype $\boldsymbol{w}_K$ one obtains an update of the form

$$\boldsymbol{w}_J^{\text{new}} = \boldsymbol{w}_J + \alpha_1 \cdot \gamma^{+} \cdot 2\Lambda(\boldsymbol{\xi}-\boldsymbol{w}_J), \tag{6}$$

$$\boldsymbol{w}_K^{\text{new}} = \boldsymbol{w}_K + \alpha_1 \cdot \gamma^{-} \cdot 2\Lambda(\boldsymbol{\xi}-\boldsymbol{w}_K). \tag{7}$$

The corresponding matrix update reads

$$\frac{\partial d_L^{\Lambda}}{\partial \Omega_{mn}} = 2[\Omega(\boldsymbol{\xi}-\boldsymbol{w}_L)]_m \cdot (\xi_n - w_{L,n}),$$

$$\frac{\partial \mu}{\partial \Omega_{mn}} = \left(\gamma^{+}\frac{\partial d_J^{\Lambda}}{\partial \Omega_{mn}} + \gamma^{-}\frac{\partial d_K^{\Lambda}}{\partial \Omega_{mn}}\right),$$

$$\Omega_{mn}^{\text{new}} = \Omega_{mn} - \alpha_2 \cdot \frac{\partial \mu}{\partial \Omega_{mn}}. \tag{8}$$

Detailed information about the algorithm, parameters, running time and complexity can be found, for example in [11,2,30,7].

The algorithm is designed for classification tasks, but it also delivers a discriminant dissimilarity measure and transformation for the specific data domain, which we use in the CBIR system to enhance its performance. We determine a discriminative three-dimensional representation of the data by applying LiRaM LVQ supervised training. We chose the target dimension three in order to compare directly with previous work [4] and because we are dealing with color representations, which are usually described in three dimensions. A further advantage of this choice is, that a visualization of the data set is also possible.

Furthermore it is possible to learn local metrics in different areas of the feature space. To this end, local matrices Ω_l are attached to the prototypes $\boldsymbol{w}_l$ in the supervised training process (see [8] for details). We refer to this modification as localized LiRaM LVQ. The distance measure changes in this case to

$$d^{\Lambda_l}(\boldsymbol{w}_l,\boldsymbol{\xi}) = (\boldsymbol{w}_l-\boldsymbol{\xi})^{\top}\Lambda_l(\boldsymbol{w}_l-\boldsymbol{\xi}), \tag{9}$$

with adaptive local, symmetric and positive semi-definite matrices Λ_l corresponding to piecewise quadratic decision boundaries. Positive semi-definite-ness and symmetry can again be guaranteed by decomposing $\Lambda_l = \Omega_l^{\top}\Omega_l$ with $\Omega_l \in \mathbb{R}^{M\times N}$ with $M \leq N$, so that the data is transformed locally by Ω_l according to the classification task. In this way the rank of the matrix Λ_l is limited by M, which also may vary for different l. Regularization schemes can be used to force the system to ensure a rank of M [31].

2.3. LiRaM LVQ settings

The results of the LiRaM LVQ algorithm display a dependence on the initial state of the matrix Ω in the training. Hence, we present results on average over several random initial configurations. For the training we employ the following cross validation procedure: The data set is split in ten disjoint subsets with approximately the same composition of classes. The union of nine subsets is used to determine the transformation matrix Ω for the vectors of the remaining subset. In this way, the matrix Ω which is applied to a given feature vector from the set is obtained without using that feature vector. This procedure is repeated ten times, once for every possible selection of the subset for which Ω is determined. In addition we repeat each training process for ten different random initializations of the LiRaM LVQ algorithm, resulting in 100 runs.

We start the matrix learning after $t_M=50$ of altogether 500 epochs and apply a learning rate schedule which has proven advantageous in many implementations of relevance learning [2,15,30]. It is of the form

$$\alpha_1(t) = \frac{\alpha_1^{\text{start}}}{1+(t-1)\Delta\alpha_1}, \quad \alpha_2(t) = \frac{\alpha_2^{\text{start}}}{1+(t-t_M)\Delta\alpha_2}. \tag{10}$$

Here, t corresponds to the current epoch, i.e. sweep through the set of training data, and α_1^{start} and α_2^{start} denote the initial learning rates for the prototypes and the matrix learning. In our experiments we chose $\alpha_1^{\text{start}} = 0.01$, $\Delta\alpha_1 = \Delta\alpha_2 = 0.0001$ and $\alpha_2^{\text{start}} = 0.001$, we do not perform an optimization of these parameters concerning the retrieval rates. In our experiments we use four prototypes (one per class) and their initial positions $\boldsymbol{w}_i(t=0)$ are determined as the mean over a random selection of 1/3 of the available feature vectors in class $c(\boldsymbol{w}_i)$ with small random deviation. Hence, prototypes are initially close to the class-conditional means in the training data, but with small deviations due to the random sampling. This has the advantage that in the case of more prototypes it is ensured that they are not initialized on exactly the same position. Relevance initialization is done by generating independent uniform random numbers $\Omega_{ij} \in [-1,1]$ and subsequent normalization, such that

$$\sum_i \Lambda_{ii} = \sum_{mn} \Omega_{mn}^2 = 1. \tag{11}$$

Performing independent runs with random initialization and subsequent normalization prevents that single features are favored by unlucky initialization. In the experiments we consider matrices $\Omega \in \mathbb{R}^{3\times 6}$, which transform the original six-dimensional feature vector into a three-dimensional space. More dimensions do not increase the performance significantly, but using less than three caused decreasing retrieval rates. Furthermore, with three dimensions we can directly compare to earlier experiments.

The localized LiRaM LVQ is trained under the same conditions and learning rate schedules, but four matrices Ω_l are adapted together with their associated prototypes $\boldsymbol{w}_l$ in the supervised training process.

For each subset D^s, $s=1,\ldots,10$, of the data set X we perform 10 runs over random initializations $i=1,\ldots,10$. For every image x_n with $n=1,\ldots,211$ from the data set we compute the correct retrieval rate by means of the k nearest neighbors within $X\backslash\{x_n\}$. Therefore, we apply for each initialization i the transformation Ω^{si} or Ω_l^{si} in the localized version, which was learned without the samples $x \in D^s$, and obtain a retrieval rate r_n^i for the query $x_n \in D^s$. Thus we get for every initialization i a mean retrieval rate $\bar{r}^i = \frac{1}{211}\sum_{n=1}^{211} r_n^i$. As an overall estimate of the performance we determine the total mean rate $r = \frac{1}{10}\sum_i^{10} \bar{r}^i$. The variability with respect to initialization is quantified by the standard deviation

$$\sigma_{\text{init}} = \left(\frac{1}{9}\sum_{i=1}^{10}(\bar{r}^i - r)^2\right)^{1/2}. \tag{12}$$

In order to quantify the variation of the data set we evaluate the mean retrieval rate of every image $\bar{r}_n = \frac{1}{10}\sum_{i=1}^{10} r_n^i$ and the corresponding standard error of mean (SEM)

$$\varepsilon_{\text{data}} = \left(\frac{1}{210}\sum_{n=1}^{211}(\bar{r}_n - r)^2\right)^{1/2} \cdot 211^{-1/2}. \tag{13}$$

With the original features there is no training process involved and $\varepsilon_{\text{data}}$ in Eq. (13) is computed simultaneously with the retrieval rate r_n of every image replacing $\bar{r}_n$.

2.4. Feature transformation obtained by LMNN

The k nearest neighbor (kNN) algorithm is a simple and intuitive method which classifies a novel feature vector by a majority vote among its k nearest neighbors in the training set. Thus, its performance depends crucially on the metric used for the identification of the neighbors. The Large Margin Nearest Neighbor (LMNN) [45] algorithm extends the kNN rule by an adaptive distance measure. The aim of the training process is that a predefined number κ of nearest neighbors (called target neighbors) belongs to the same class like the example data with high probability. Simultaneously, samples of different classes should be separated by a large margin. The corresponding optimization problem is convex and the global optimum can be found by means of semi-definite programming [45]. The computational effort depends crucially on the parameter κ. The LMNN algorithm provides a discriminative distance measure for the kNN classifier corresponding to $d_\Psi(\xi_i,\xi_j) = (\xi_i-\xi_j)^\top \Psi(\xi_i-\xi_j)$. Here, the matrix $\Psi \in \mathbb{R}^{M\times N}$ denotes the counterpart of Ω in LiRaM LVQ.

The training procedure has two steps. The first step identifies a set of κ similarly labeled target neighbors for each input ξ_i. The second step adapts the Mahalanobis distance metric so that these target neighbors are closer to ξ_i than differently labeled inputs. The semidefinite optimization in LMNN classification arises from an objective function which balances two terms. The first term penalizes large distances between inputs and their target neighbors. The second term penalizes small distances between differently labeled inputs. The terms in the objective function can be specified with further notation. Let $y_{ij} \in \{0,1\}$ indicate whether the inputs ξ_i and ξ_j have the same class label. The notation $j \rightsquigarrow i$ indicates that ξ_j is a target neighbor of ξ_i. Also, let $x_{ijl} \geq 0$ denote the amount by which a differently labeled input ξ_l invades the perimeter around input ξ_i defined by its target neighbor ξ_j. The Mahalanobis distance metric Ψ is obtained by solving the semidefinite program shown in Table 1. The constant μ defines the trade-off between the two terms in the objective function. The constraints of type (a) favor inputs ξ_i closer to their κ target neighbors ξ_j then to any other differently labeled input ξ_l. When differently labeled ξ_l invade the local neighborhood a positive slack variable x_{ijl} is generated. This is penalized in the second term of the objective function. Constraints of type (b) enforce non-negativity of the slack variables and constraint (c) enforces positive semi-definiteness of Ψ. Noting that the squared Mahalanobis distances are linear in the matrix Ψ, the above optimization is easily recognized as a semidefinite problem.

Table 1
Semidefinite optimization problem in LMNN.

Minimize $\sum_{j\rightsquigarrow i}[d_\Psi(\xi_i,\xi_j) + \mu\sum_l(1-y_{il}x_{ijl})]$
subject to:
(a) $d_\Psi(\xi_i,\xi_l) - d_\Psi(\xi_i,\xi_j) \geq 1 - x_{ijl}$
(b) $x_{ijl} \geq 0$
(c) $\Psi \succcurlyeq 0$

The results presented in the following section were produced with the first code available at www.cse.wustl.edu/~kilian/code/code.html (last visited September 2010) [45] using default parameters except for the number of target neighbors κ, which varies in our experiments from 1 to 25 and the initial matrix $\Psi \in \mathbb{R}^{3\times 6}$ with elements randomly drawn from the interval $[-1,1]$. For a fair comparison, LMNN and LiRam LVQ are applied to the same subsets D^s of training data and performance is evaluated on the same footing.

2.5. Canonical representations

Note that the transformation matrix Ω obtained by LiRaM LVQ and Ψ in LMNN are not uniquely determined: For instance, the distance measure is invariant under rotations in the feature space. Thus, the training process can yield different transformation matrices Ω depending on the (random) initialization of the training process. We identify uniquely defined transformations $\hat{\Omega}$ and $\hat{\Psi}$ by decomposing $\Lambda = \Omega^\top\Omega$ and $\Upsilon = \Psi^\top\Psi$ in a canonical way: we determine the eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_M$ corresponding to the M (ordered) non-zero eigenvalues of Λ or Υ, $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_M$ and define $\hat{\Omega}$ or $\hat{\Psi}$ as follows:

$$\hat{\Psi}, \hat{\Omega} = ([\sqrt{\lambda_1}\boldsymbol{v}_1, \sqrt{\lambda_2}\,\boldsymbol{v}_2, \ldots, \sqrt{\lambda_M}\boldsymbol{v}_M])^\top \in \mathbb{R}^{M\times N}. \tag{14}$$

This canonical representation does not alter the retrieval system and it allows direct comparison of the transformations $\hat{\Omega}$ and $\hat{\Psi}$.

It is not obvious how to extend the LMNN scheme for a comparison with the use of local matrices Ω_l like in the LiRaM LVQ. We will discuss the localized matrices in terms of the achieved retrieval performance and show the mean canonical representations.

2.6. Retrieval test

As a performance measure for CBIR we use the average correct retrieval rate, also referred to as precision. It is defined as the percentage of k nearest neighbors that belong to the same category as a query image. We determine for each image its k nearest neighbors in the entire data set using the Euclidean distance measure. For comparison, we do this both in the original feature space ξ and in the transformed feature space $\tilde{\xi} = L\xi$ with $L \in \{\Omega, \Psi\}$. Note that in our evaluation for a given query image, the

Table 2
Color representations.

Color space	Chosen for
RGB	Widespread use
Normalized RGB	Invariance (under certain assumptions) to changes of surface orientation with respect to the light source [35]
TSL	Successful application in skin detection [40]
CIE-XYZ	Role as the basis for CIE-Lab and CIE-Luv
CIE-Lab	Perceptual relevance and relation to melanin and hemoglobin [38]
CIE-Luv & CIE-Lch	Perceptual relevance
YCrCb	Simplicity and explicit separation of luminance and chrominance components [25,46] and popularity in skin detection applications [18]

transformation matrices Ω, Ψ and Ω_l have been determined from subsets which do not contain the query.

Using the localized LiRaM LVQ approach the training process optimizes l localized transformations Ω_l corresponding to the classification task. We involve this information by projecting every feature vector $\boldsymbol{\xi}$ with the transformation Ω_l corresponding to the nearest prototype $\boldsymbol{w}_l$ with $d^{\Lambda_l}(\boldsymbol{w}_l,\boldsymbol{\xi}) < d^{\Lambda_k}(\boldsymbol{w}_k,\boldsymbol{\xi}) \forall l \neq k$ resulting in local linear projections for different areas of the feature space.

Section 3 presents and compares the resulting retrieval rates as average over all images. Furthermore, the standard error of the performance with the actual query image and its dependence on the initialization of LiRaM LVQ are discussed.

2.7. Color spaces

We explore the retrieval rates for eight different color representations separately. The different color spaces vary, as already mentioned, with respect to their usefulness in different applications. Possible motivations for the choice of a particular color space are summarized in Table 2.

Despite the potential difficulty rising from the cyclic representation of the "Hue" component of the TSL color space and its relatives HSV and HSL, for completeness, we investigate its behavior for our application task in terms of one example, namely TSL.

3. Results

3.1. Retrieval rate

In this Section we summarize the retrieval results for the different color representations using transformed features from LMNN, global and localized LiRaM LVQ. We compare them with those obtained in the original feature spaces and with the difference features from [4] obtained with the transformation A:

$$A = \begin{pmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{pmatrix}. \quad (15)$$

The overall mean rates r obtained with LiRaM LVQ and $\Omega \in \mathbb{R}^{3\times 6}$ are displayed in Fig. 5 for each color space as a function of the number k, i.e. the number of pictures the CBIR system returns to the user. The best correct retrieval rates for this algorithm are achieved with the color spaces YCrCb (82.3%), CIE-Lab (82.2%), CIE-Lch (81.1%), CIE-Luv (81.0%) and RGB (80.7%) where the numbers correspond to the example case $k=11$. All other color representations yield by far lower performances with rates between 68.7% and 75.0%. We chose the example case of 11 returned images for the quantitative analysis to be able to compare to earlier studies [4] and because it seems a reasonable large number suggested by the doctor. Of course the system is able to return as many similar images as the data base contains and the user wishes to see.

Fig. 6 shows a comparison of the correct retrieval rates based on the original features (red lines), the difference features from [4] (green lines) and the transformed data (blue and black lines) as a function of the neighborhood size k of the retrieval system. The gray shaded areas mark the SEM $\varepsilon_{\text{data}}$, while the blue shaded area corresponds to σ_{init} of the LiRaM LVQ. Note that the latter is, of course, absent in the results based on original features and difference features, as no training process is involved and also absent in the results coming from LMNN, because it finds the global optimum for a given parameter set, independent of the initial state. The variation due to initialization of the localized LiRaM LVQ is not displayed; it is comparable to the variation in the global version. We set the parameter κ of the LMNN approach equal to the neighborhood k of the retrieval system and, in addition, we consider $\kappa = 25$. The latter is close to the size of the smallest class in the data set, "blue" (c), with 29 examples. For $\kappa = 25$ the retrieval performances of LMNN and LiRaM LVQ are comparable which is also reflected in the fact that the obtained matrices $\hat{\Omega}$ and $\hat{\Psi}$ are very similar, cf. Figs. 7 and 8. Smaller values for κ reduce the computational effort of the optimization at the expense of performance.

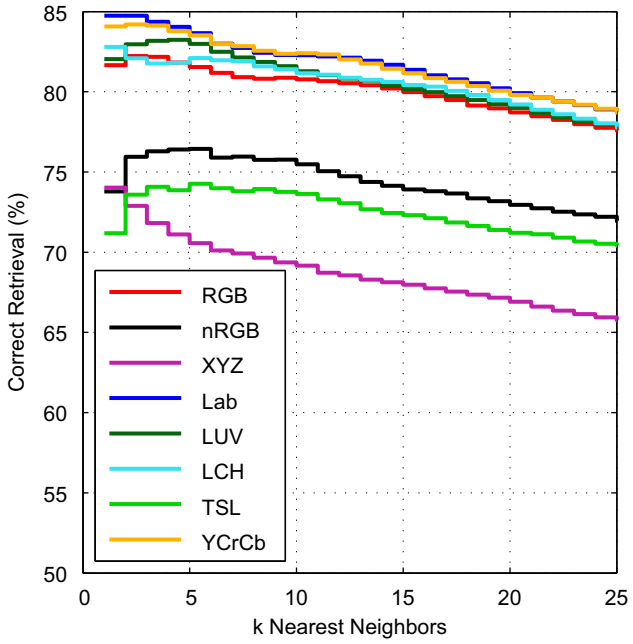

Fig. 5. Mean correct retrieval rates obtained with the LiRaM LVQ transformed data as a function of the number k of retrieved images for eight color spaces.

Localized LiRaM LVQ achieves the best correct retrieval rate for the most suitable color spaces: Lab and YCrCb. However, the performance boost compared to the other methods is only moderate. In TSL, localized LiRaM LVQ is even outperformed by the simpler techniques based on global measures. These findings suggest that the latter already extract the most important information from the original color features. Furthermore, TSL is cyclic represented by the angle of color components, which may cause instabilities for naive distance computation. We suggest the performance drop of the

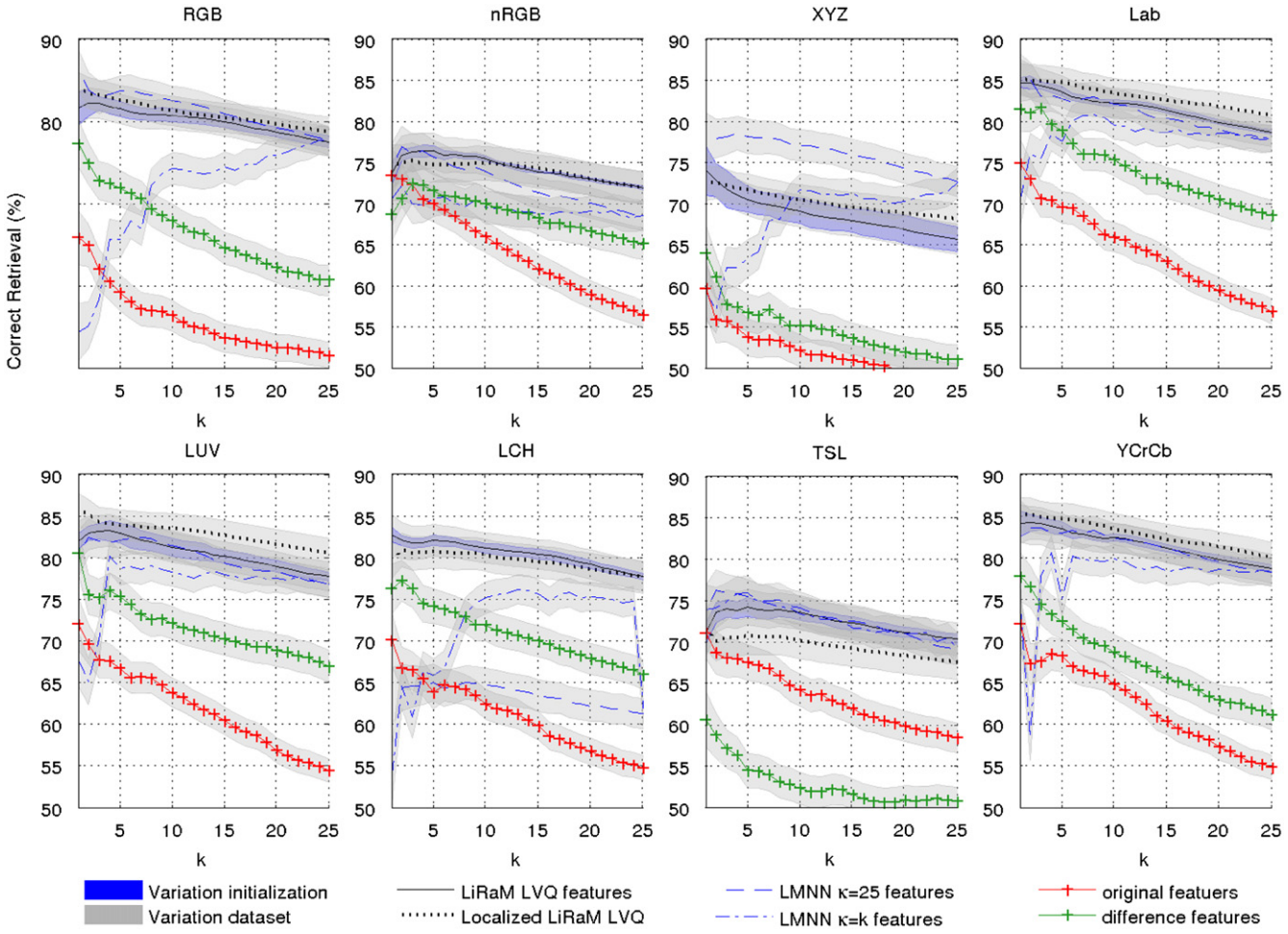

Fig. 6. Comparison of correct retrieval rates in dependence on the number of nearest neighbors k for each color space. The red lines denote the mean retrieval rates on the original feature space, the green line stands for the difference features from [4], whereas the blue and black lines shows the mean results on the transformed feature spaces. The blue shaded areas indicates the standard deviation due to the random initializations in LiRaM LVQ.

difference features in comparison to the use of the original features is a consequence of the Hue representation in TSL and its relatives HSL and HSU were we observed the same effect. However, the adaptive distance is able to compensate for this effect and still yields a boost of performance also in these color spaces.

In most of the color spaces, including RGB, the LiRaM LVQ result is not very sensitive to initialization, as indicated by relatively small standard deviations $\sigma_{\text{init}} \leq 2\%$. The XYZ color representation display the largest dependence on initialization with $\sigma_{\text{init}} > 2.7\%$. The variation with the data set is approximately the same in original and transformed feature spaces. This variability is not an effect of the LiRaM LVQ training but is characteristic of the data set itself.

In the case of the LMNN optimization, we observe that the use of an adaptive transformation increases the mean retrieval rate r significantly for all color spaces, for every choice of k and appropriate κ. The best results are obtained with CIE-Lab ($72\% < r < 85\%$) and YCrCb ($72\% < r < 84\%$). It is interesting to note that the popular RGB representation exhibits comparable performance ($70\% < r < 82\%$) in the transformed feature space.

Thus, we achieve an improvement between 10% and 27% when employing an adaptive linear transformation of features.

3.2. Recommended transformations

Here we inspect the favorable transformations of the feature space as obtained by LiRaM LVQ and LMNN. We focus on RGB as the by far most frequently used color space and on CIE-Lab because of its excellent retrieval performance.

3.2.1. Global transformations

We observe that the obtained distance measure represented by Λ depends only weakly on the initialization of LiRaM LVQ. However, a continuum of matrices Ω satisfies $\Omega^\top \Omega = \Lambda$ and, in this sense, the actual outcome Ω of the training process can vary widely. Thus, the canonical representation of $\hat{\Omega}$ is averaged over all training runs. The mean transformation is explicitly given for RGB in Eq. (16) and visualized in Fig. 7. The standard deviation concerning the random initialization of each component lies between 0.01 to 0.03 for $\hat{\Omega}_{\text{RGB}}$. Each row of the matrix defines a new feature as a linear combination of the original six features:

$$\hat{\Omega}_{\text{RGB}} = \begin{pmatrix} 0.139 & -0.192 & 0.093 & -0.320 & 0.662 & -0.469 \\ 0.127 & -0.082 & -0.112 & -0.167 & 0.080 & 0.276 \\ 0.036 & -0.064 & 0.108 & -0.047 & -0.063 & -0.002 \end{pmatrix}. \tag{16}$$

We observe, that the absolute weights corresponding to skin lesions (columns 4,5,6) are typically 1–2 times larger than the coefficients assigned to the healthy skin features (columns 1, 2, 3). In general, the corresponding coefficients for lesion and healthy skin features are of opposite sign. Hence, the transformed features correspond to weighted differences of the lesion and healthy skin

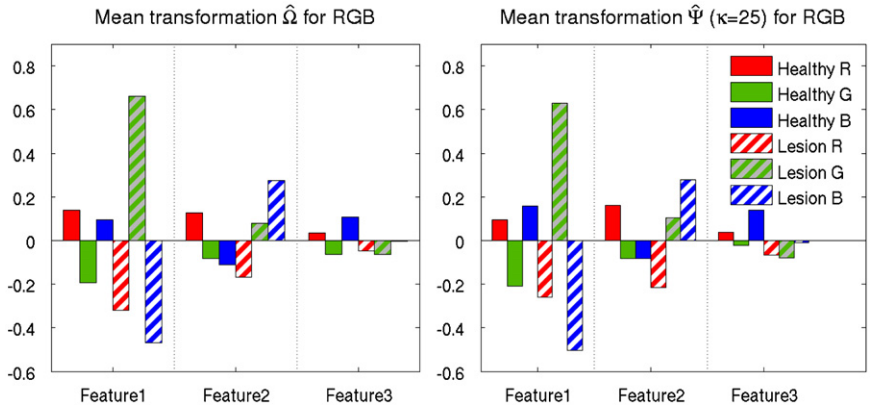

Fig. 7. Recommendation for the transformation in RGB: (right) multipliers that define the new features as linear combinations of the original features earned from LiRaM LVQ and (left) multipliers earned from LMNN with $\kappa = 25$.

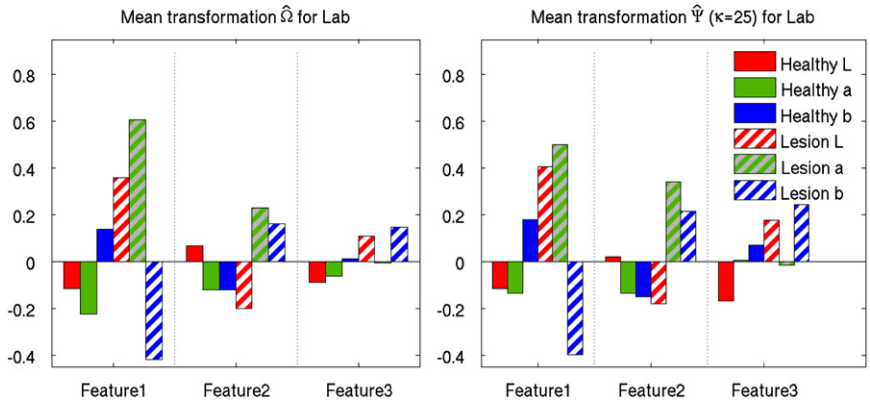

Fig. 8. Recommendation for the transformation in CIE-Lab: (top) multipliers that define the new features as linear combinations of the original features earned from LiRaM LVQ and (left) multipliers earned from LMNN with $\kappa = 25$.

color values. Eq. (17) denotes explicitly the mean transformation $\hat{\Omega}_{\text{Lab}}$ for CIE-Lab; it is visualized in Fig. 8. The above discussed properties of Ω_{RGB} persist also in the transformation of CIE-Lab feature vectors. The standard deviations for the mean transformation vary from 0.01 and 0.06 for the random initializations:

$$\hat{\Omega}_{\text{Lab}} = \begin{pmatrix} -0.115 & -0.225 & 0.140 & 0.358 & 0.606 & -0.418 \\ 0.069 & -0.120 & -0.120 & -0.200 & 0.231 & 0.164 \\ -0.087 & -0.063 & 0.011 & 0.109 & -0.006 & 0.147 \end{pmatrix}. \tag{17}$$

The resulting 3D visualizations of the data set with the mean canonical transformations $\hat{\Omega}$ using the RGB and LAB color representation are shown in Fig. 9. It can be seen that the classes for "white", "red" and "brown" skin cancer build a nicely separable data cloud respectively, whereas the class "blue" lays between the others and overlaps. With more training samples especially of the difficult class the data set might be even better separable by supervised adaptive dissimilarity learning.

3.2.2. Local transformations

Also with the localized matrices the above discussed properties persist. For the local feature transformation the prototypes are necessary and define the area of the original feature space, where their transformation is valid. So the samples are transformed with the transformation attached to the nearest prototype $\boldsymbol{w}_j$:

$$\tilde{\boldsymbol{x}} = \Omega_j \boldsymbol{x} \quad \text{with } d^{\Lambda_j}(\boldsymbol{w}_j, \boldsymbol{x}) = \min_k d^{\Lambda_k}(\boldsymbol{w}_k, \boldsymbol{x}). \tag{18}$$

The mean canonical representations of the local matrices for RGB are shown in Fig. 10. Note that the definition in Eq. (18) is only valid in the neighborhood of the corresponding prototype. At the borders of the Voronoi cell of each prototype this definition may be inappropriate. In general it is possible to combine the local linear patches in a global nonlinear way by charting [6,5] or Local Linear Coordination (LLC) [39]. It can be seen that some class-wise transformations seems to be already well discriminating with one or two features, for example the matrices for the "brown" and "red" class of skin lesions. However, for the class of white and bluish appearing skin lesions also the third feature shows a contribution to the transformation. It would have been possible to have class-wise different target spaces for two and one dimension in respective transformations, but for reasons of consistency and for comparison purpose we chose the target dimension to be the same for every class.

In summary, our findings support the basic idea of using differences of color features presented in [4]. We have shown, however, that generalizing this concept by introducing adaptive coefficients improves the retrieval performance significantly for this supervised problem.

4. Summary and conclusion

In this paper show the usefulness of adaptive distances and corresponding feature space transformations on an example real world application. We observe that CBIR on color is a powerful tool for analysis of dermatological image databases. Previously unnoticed color similarities may give new insight into the correlations

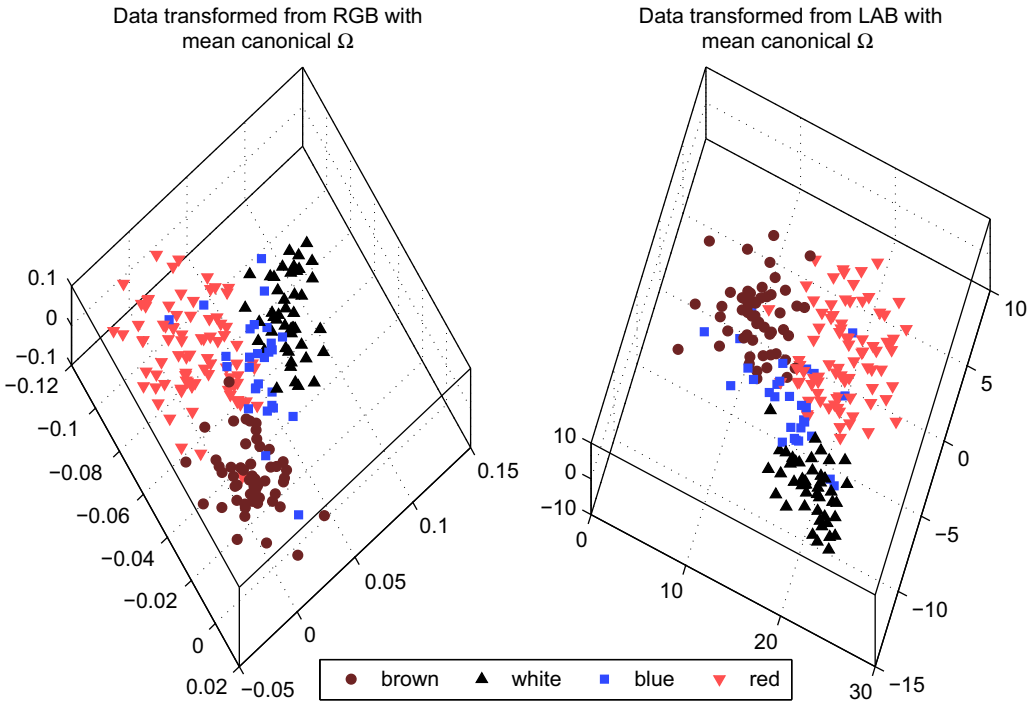

Fig. 9. The resulting 3D visualizations of the skin cancer data set transformed from the RGB and LAB color space with $\hat{\Omega}_{\mathrm{RGB}}$ Eq. (16) (left panel) and $\hat{\Omega}_{\mathrm{Lab}}$ Eq. (17) (right panel).

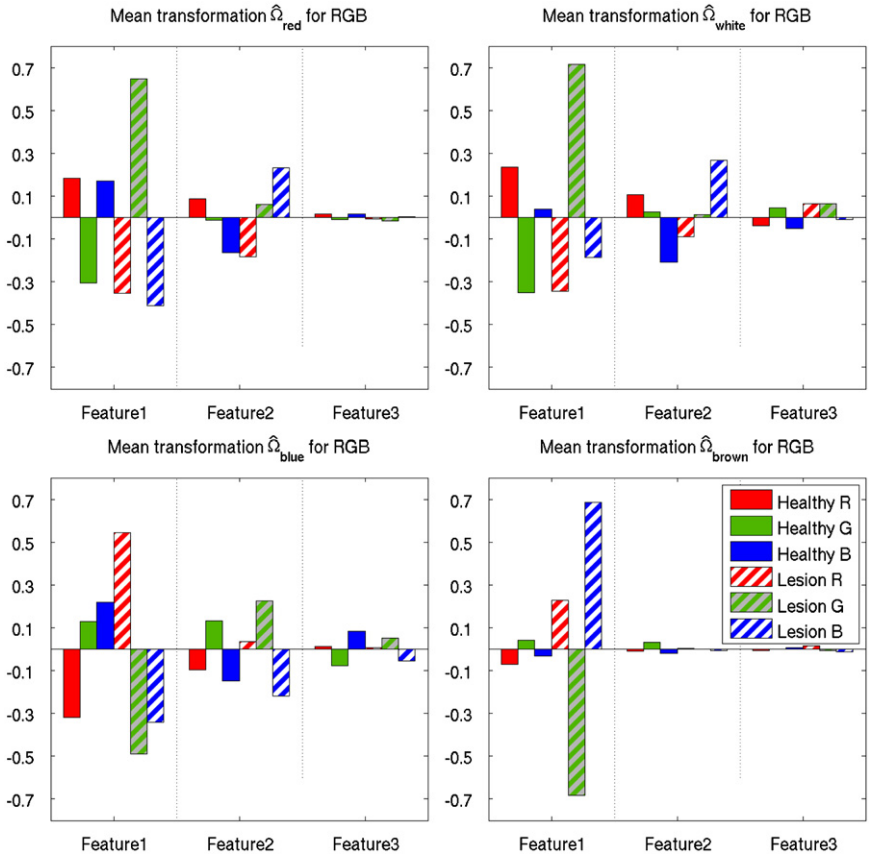

Fig. 10. Local matrices for RGB corresponding to one prototype of each class.

between and within various skin diseases. We introduce discriminative color descriptors which are obtained by LiRaM LVQ and LMNN during supervised training, and we compare and evaluate their performance for CBIR of dermatological images. Starting from a 6D vector representation of images, we define three new features as linear combinations of the original six color components of

healthy and lesion skin. The linear combinations are determined by LiRaM LVQ in a training process which is guided by classification performance and yields a discriminative representation of the feature space. With new features we achieve considerable improvement of retrieval results in all eight color spaces that we studied. In the five best color spaces (YCrCb, CIE-Lab, CIE-Lch, CIE-Luv and RGB) the increase of the correct retrieval rate is between 10% and 27% in the range of $k=1$–25 retrieved images in comparison to earlier studies. We conclude that adaptive dissimilarity learning is favorable independent of the choice of the actual color space. The user may decide according to his personal preference which color representation is most suitable.

The use of LMNN seems natural, since the retrieval is based on a kNN approach. However, our investigation shows that the LiRaM LVQ approach outperforms LMNN if the latter takes only a relatively small number κ of neighbors into account in the training process. For larger κ the obtained metric becomes very similar to that of LiRaM LVQ and, consequently, the retrieval performances are comparable. The computational effort for LiRaM LVQ training is typically lower than that of the LMNN optimization which grows with κ. An important advantage of the LVQ approach is its potential with respect to extensions. As shown, for example, local metrics can be attached to the prototypes which are responsible for different areas of the original feature space. In the most favorable color spaces, localized LiRaM LVQ increased the retrieval rates even further.

We conclude that LiRaM LVQ is an efficient technique for the extraction of highly discriminative color features for CBIR of dermatological images. With this approach, we obtain high mean correct retrieval rates of between 84% for $k=1$ and 79% for $k=25$ retrieved images in the five best color spaces. For two of the color spaces, RGB and CIE-Lab, we discuss in detail the canonical linear transformations of the original six color components to three new features and showed their superiority to recently introduced approaches.

Obviously, several important extensions are possible. For instance, the automatic detection of regions of interest or the integration of shape information should be relevant in practical applications. Forthcoming studies should address, among other modifications, the use of extended original feature spaces which include, for instance, shape information.

Acknowledgements

The authors thank Piet Toonder of the UMCG for making available the image data set used in this study. This work was supported by the "Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)" under project code 612.066.620.

References

[1] A. Blum, H. Luedtke, U. Ellwanger, R. Schwabe, G. Rassner, C. Garbe, Digital image analysis for diagnosis of cutaneous melanoma. Development of a highly effective computer algorithm based on analysis of 837 melanocytic lesions, British Journal of Dermatology 151 (5) (2004) 1029–1038.

[2] T. Bojer, B. Hammer, D. Schunk, K.T. von Toschanowitz, Relevance determination in learning vector quantization, in: M. Verleysen (Ed.), Proceedings of European Symposium on Artificial Neural Networks (ESANN), 2001, pp. 271–276.

[3] J.L. Bolognia, J.L. Jorizzo, R.P. Rapini, Dermatology, second ed., Mosby, 2007.

[4] H.H.W.J. Bosman, N. Petkov, M.F. Jonkman, Comparison of color representations for content based image retrieval in dermatology, Skin Research and Technology. 16 (2010) 109–113.

[5] M. Brand, Charting a manifold, Technical Report 15, Mitsubishi Electric Research Laboratories (MERL), 2003, URL: 〈http://www.merl.com/publications/TR2003-013/〉.

[6] K. Bunte, B. Hammer, P. Schneider, M. Biehl, Nonlinear discriminative data visualization, in: M. Verleysen (Ed.), Proceedings of European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium, April 2009, pp. 65–70.

[7] K. Bunte, B. Hammer, A. Wismüller, M. Biehl, Adaptive local dissimilarity measures for discriminative dimension reduction of labeled data, Neurocomputing 73 (7–9) (2010) 1074–1092.

[8] K. Bunte, P. Schneider, B. Hammer, F.-M. Schleif, T. Villmann, M. Biehl, Discriminative visualization by limited rank matrix learning, Technical Report MLR-03-2008, University Leipzig, 2008, URL: 〈http://www.uni-leipzig.de/~compint/mlr/mlr_03_2008.pdf〉.

[9] D. Chai, A. Bouzerdoum, A Bayesian approach to skin color classification in YCbCr color space, in: Proceedings IEEE Region Ten Conference (TENCON'2000), vol. 2, 2000, pp. 421–424.

[10] Y. Cheng, R. Swamisai, S.E. Umbaugh, R.H. Moss, W.V.W. Stoecker, S. Teegala, S.K. Srinivasan, Skin lesion classification using relative color features, Skin Research and Technology 14 (2008) 53–64.

[11] K. Crammer, R. Gilad-bachrach, A. Navot, N. Tishby, Margin analysis of the LVQ algorithm, in: Advances in Neural Information Processing Systems 2002, MIT Press, 2002, pp. 462–469.

[12] R. Datta, J. Li, J.Z. Wang, Content-based image retrieval: approaches and trends of the new age, in: Proceedings of Seventh ACM SIGMM International Workshop on Multimedia Information Retrieval, 2005, pp. 253–262.

[13] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., Wiley-Interscience, November 2004.

[14] C.D. Felice, M.L. Flori, M. Pellegrino, P. Toti, E. Stanghellini, A. Molinu, P. Tosi, F. Bagnoli, Predictive value of skin color for illness severity in the high-risk newborn, Pediatric Research 51 (1) (2002) 100–105.

[15] B. Hammer, T. Villmann, Generalized relevance learning vector quantization, Neural Networks 15 (8–9) (2002) 1059–1068.

[16] K. Hoffmann, T. Gambichler, A. Rick, Diagnostic and neural analysis of skin cancer DANAOS). A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy, British Journal of Dermatology 149 (4) (2003) 801–809.

[17] A.K. Jain, A. Vailaya, Image retrieval using color and shape, Pattern Recognition 29 (8) (1996) 1233–1244.

[18] P. Kakumanu, S. Makrogiannis, N. Bourbakis, A survey of skin-color modeling and detection methods, Pattern Recognition 40 (2007) 1106–1122.

[19] R. Kjeldsen, J. Kender, Finding skin in color images, in: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, Killington, Vermont, 1996, pp. 312–317.

[20] T. Kohonen, Self-Organizing Maps, third ed., Springer, Berlin, Heidelberg, New York, 2001.

[21] T. Lehmann, M. Güld, C. Thies, B. Fischer, K. Spitzer, D. Keysers, H. Ney, M. Kohnen, H. Schubert, B. Wein, Content-based image retrieval in medical applications, Methods of Information in Medicine 43 (4) (2004) 354–361.

[22] R. Min, H.D. Cheng, Effective image retrieval using dominant color descriptor and fuzzy support vector machine, Pattern Recognition 42 (1) (2009) 147–157.

[23] H. Müller, N. Michoux, D. Bandon, A. Geissbuhler, A review of content-based image retrieval systems in medical applications—clinical benefits and future directions, International Journal of Medical Informatics 73 (2004) 1–23.

[24] G. Pass, R. Zabih, J. Miller, Comparing images using color coherence vectors, in: Proceedings of ACM Multimedia 96, Boston, MA, November 1996, pp. 65–73.

[25] S.L. Phung, A. Bouzerdoum, D. Andchai, A novel skin color model in YCbCr color space and its application to human face detection, in: IEEE International Conference on Image Processing, vol. 1, 2002, pp. 289–292.

[26] G.G.F. Roli, Bayesian relevance feedback for content-based image retrieval, Pattern Recognition 37 (2004) 1499–1508.

[27] A.S. Sato, K. Yamada, Generalized learning vector quantization, Advances in Neural Information Processing Systems 8 (1996) 423–429.

[28] P. Schmid-Saugeona, J. Guillodb, J.P. Thirana, Towards a computer-aided diagnosis system for pigmented skin lesions, Computerized Medical Imaging and Graphics 27 (1) (2003) 65–78.

[29] P. Schneider, M. Biehl, B. Hammer, Relevance matrices in LVQ, in: M. Verleysen (Ed.), Proceedings of European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium, 2007, pp. 37–42.

[30] P. Schneider, M. Biehl, B. Hammer, Adaptive relevance matrices in learning vector quantization, Neural Computation 21 (12) (2009) 3532–3561.

[31] P. Schneider, K. Bunte, B. Hammer, M. Biehl, Regularization in matrix relevance learning, IEEE Transactions on Neural Networks 21 (5) (2010) 831–840.

[32] P. Schneider, K. Bunte, B. Hammer, T. Villmann, M. Biehl, Regularization in matrix relevance learning, Technical Report, MLR-02-2008, University Leipzig, 2008, URL: 〈http://www.uni-leipzig.de/~compint/mlr/mlr_02_2008.pdf〉.

[33] J.L. Shih, L.H. Chen, Colour image retrieval based on primitives of colour moments, Vision, Image and Signal Processing, IEE Proceedings 149 (December) (2002) 370–376.

[34] M.C. Shin, K.I. Chang, L.V. Tsap, Does colorspace transformation make any difference on skin detection?, in: WACV '02: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision, IEEE Computer Society, Washington, DC, USA, 2002, pp. 275–279.

[35] W. Skarbek, A. Koschan, Colour image segmentation—a survey, Technical Report 32, Technical University of Berlin, Department of Computer Science, October 1994, URL: 〈citeseer.ist.psu.edu/skarbek94colour.html〉.

[36] A. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, Content-based image retrieval at the end of the early years, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (12) (2000) 1349–1380.

[37] K. Sobottka, I. Pitas, Segmentation and tracking of faces in color images, in: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, Killington, Vermont, 1996, pp. 236–241.

[38] H. Takiwaki, Measurement of skin color: practical application and theoretical considerations, The Journal of Medical Investigation 44 (3-4) (1998) 121–126.
[39] Y. Teh, S. Roweis, Automatic alignment of local representations, Advances in Neural Information Processing Systems 15 (2003) 841–848 URL: 〈citeseer.ist.psu.edu/teh02automatic.html〉.
[40] J.C. Terrillon, S. Akamatsu, Comparative performance of different chrominance spaces for color segmentation and detection of human faces in complex scene images, in: Proceedings of the Vision Interface, May 1999, pp. 180–197.
[41] R.d.S. Torres, A.X. Falc ao, M.A. Gonçalves, Jo.a.P. Papa, B. Zhang, W. Fan, E.A. Fox, A genetic programming framework for content-based image retrieval, Pattern Recognition 42 (2) (2009) 283–292.
[42] S.E. Umbaugh, R.H. Moss, W.V. Stoecker, An automatic color segmentation algorithm with application to identification of skin lesion borders, Computerized Medical Imaging and Graphics 16 (1992) 227–235.
[43] V.V. Vassili, V. Sazonov, A. Andreeva, A survey on pixel-based skin color detection techniques, in: Graphicon2003, in: 13th International Conference on Computer Graphics and Vision, September 2003, pp. 85–92.
[44] H. Voigt, R. Classen, Computer vision and digital imaging technology in melanoma detection, Seminars in Oncology 29 (4) (2002) 308–327.
[45] K.Q. Weinberger, J. Blitzer, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, Advances in Neural Information Processing Systems 18 (2006) 1473–1480.
[46] B.D. Zarit, B.J. Super, F.K.H. Andquek, Comparison of five color models in skin pixel classification, in: International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems, 1999, pp. 58–63.

Kerstin Bunte graduated at the Faculty of Technology at the University of Bielefeld, Germany, and joined the Institute of Mathematics and Computing Science of the University of Groningen, The Netherlands, in September 2007. Her recent work has focused on Machine Learning techniques, especially Learning Vector Quantisation and their usability in the field of image processing, dimension reduction and visualization. Further information can be obtained from http://www.cs.rug.nl/~kbunte/

Michael Biehl received a Ph.D. in Theoretical Physics from the University of Giessen, Germany, in 1992 and the venia legendi in Theoretical Physics from the University of Würzburg, Germany, in 1996. He is currently Associate Professor with Tenure in Computing Science at the University of Groningen, The Netherlands. His main research interest is in the theory, modelling and application of Machine Learning techniques. He is furthermore active in the modelling and simulation of complex physical systems. He has co-authored more than 100 publications in international journals and conferences; preprint versions and further information can be obtained from http://www.cs.rug.nl/~biehl/

Marcel F. Jonkman received the M.D. and Ph.D. degree in Medicine from the University of Groningen, Groningen, the Netherlands. He is professor of Dermatology and chair of the department of Dermatology of the University Medical Center Groningen, the Netherlands. Previously he held a fellow position at Jefferson Medical University, Philadelphia (as Royal Dutch Academy of Sciences scholar).

He is the author of one book, and has authored over 130 scientific papers. His current research includes bullous diseases, genetic skin diseases and creating education programs for dermatologists in training. Dr. Jonkman is a member of the editorial boards of the European Journal of Dermatology, and the Journal of Dermatological Science.

Nicolai Petkov received the Dr.sc.techn. degree in Computer Engineering (Informationstechnik) from Dresden University of Technology, Dresden, Germany. He is professor of computer science and head of the Intelligent Systems group of the Institute of Mathematics and Computing Science of the University of Groningen, the Netherlands. He is the author of two monographs and coauthor of another book on parallel computing, holds four patents and has authored over 100 scientific papers. His current research is in image processing, computer vision and pattern recognition, and includes computer simulations of the visual system of the brain, computer applications in health care and life sciences and creating computer programs for artistic expression. Dr. Petkov is a member of the editorial boards of several journals.